RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

IFN Working Paper No. 1346, 2020

The Effect of Water Filtration on Cholera Mortality

Daniel Knutsson

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

THE EFFECT OF WATER FILTRATION ON CHOLERA MORTALITY*

Daniel Knutsson

June 26, 2020

Abstract

There is debate among researchers regarding the importance of water filtration in reducing mortality during the epidemiological transition. However, there is limited research on how water filtration affected cholera mortality during the second half of the 19th century. Using historical microdata, this paper provides new evidence on the importance of water filtration in reducing cholera mortality during an outbreak. The results show that access to filtered water protected almost completely against cholera mortality. Water filtration could thereby have contributed more to mortality decline than what has previously been documented.

Keywords: Water; Piped water; Filtered water; Cholera; Mortality; Public health
JEL classifications: I11; I18; N3; H51

Corresponding author:
Daniel Knutsson
Box 55665
SE-102 15
Stockholm, Sweden
Email: daniel.knutsson@ifn.se
Phone: +46722005998

*This research has benefited from financial support from the Swedish Research Council. I gratefully acknowledge comments from Per Pettersson-Lidbom, Andreas Madestam, Björn Tyrefors, and seminar participants at IFN (SEHL). This paper is a heavily revised version of a previous unpublished working paper, Knutsson (2017).

1 Introduction

During the 19th century, the bacterium Vibrio cholerae claimed millions of lives in Europe through several pandemic outbreaks (Pollitzer, 1954). Cholera was one of the most feared epidemic diseases at the time due to its rapid spread and high mortality among those infected. Although British physician John Snow had already argued in the 1850s that cholera could spread through drinking water, it would take many decades for his ideas to become widespread (Snow, 1849).[1] Although Europe remained ignorant regarding effective cholera prevention, cholera mortality started to decline, and by the turn of the 20th century, reoccurring epidemic cholera outbreaks were no more.

This paper provides direct evidence on the link between cholera mortality and access to filtered water. I evaluate whether water filtration technology in the mid-19th century provided effective protection against the disease. Water filters started to be used in Europe in the 1830s and became increasingly popular in the large industrializing cities of Western Europe. As cholera is one of few epidemic diseases that is almost exclusively transmitted through ingestion of contaminated water, water filtration is a plausible candidate explanation for the observed decline in mortality from cholera during the late 19th century (Ewald, 1991).

There are several challenges associated with studying epidemic diseases and preventive measures that this paper is able to overcome. First, and most trivial, to study cholera, researchers have to observe cholera in their data.[2] At the turn of the 20th century, cholera was close to eradicated in the Western world, making it impossible for research on this time period to measure and quantify the effect of clean water on cholera. In this paper, by contrast, the water filtration plant under study was constructed during a cholera pandemic and at a time when cholera was an imminent threat to the lives of urban residents in Europe.

However, even when cholera is present, analyzing the protective role of filtered

[1] For estimated cholera mortality during the 19th century, see table 1.

[2] Several papers study an endemic waterborne disease, typhoid fever. Although clean water technologies consistently reduce mortality from this disease, they contribute very little to the total mortality decline because the mortality rate from typhoid fever was low in most settings under study (Cutler and Miller, 2005).

water is challenging due to the data requirements. As the timing and intensity of a cholera outbreak is hard to predict, using more aggregated and hence geographically distant units of observation (such as states, cities or municipalities) to study this question is undesirable because control units may not provide valid counterfactual exposures to the treatment group. Furthermore, externalities from treated units could affect disease exposure in control units because epidemic diseases travel between different locations. If cities with clean water prevent diseases from spreading to untreated cities, treatment and control units are not well defined (Imbens and Rubin, 2015, pp.10). Using within-city variation is thus preferable.

Cholera outbreaks normally have a rapid course, and an epidemic often starts and ends in just a few months (Koch, 1894). High-frequency data are thus required to precisely map cholera mortality to filtered water access. When using yearly data, the kind of data normally available to researchers, the cholera shock would be more difficult to separate from other shocks. Moreover, a short epidemic can avoid selective sorting of residents into different houses over time as a consequence of the shock. The cholera outbreak studied in this paper lasted for four months. By analyzing this shock using address-level information on access to filtered water and mortality at a monthly frequency, I can avoid many of the challenges involved with studying cholera.

The water filtration technology in Stockholm was, at the time, a state-of-the-art slow sand filter that was delivered through a piped distribution network. I can study this technology in itself because there was no simultaneous construction of sewerage systems, something that is otherwise common (see, e.g., Alsan and Goldin, 2019). Treatment heterogeneity is further limited by only one, and the same, technology being provided to all treated houses. This research thereby focus attention on a specific technology and its role in preventing cholera mortality. Comparing different cities or municipalities is more complicated, as the water technologies under study will differ in a wide range of attributes depending on the water sources, water cleaning technologies, and distribution systems.[3]

[3] Estimating standard errors of several treatments in difference-in-differences designs is challenging due, inter alia, to the small-sample issues that often follow from having few groups in each treatment arm (Cameron and Miller, 2015; MacKinnon and Webb, 2017).

Interest in how public health efforts can explain historical declines in mortality has a long history in the social sciences (McKeown and Brown, 1955).[4] At present, an extensive literature is devoted to investigating how water, sewerage, and sanitation technologies can explain the mortality decline observed in several historical settings (see, e.g., Cutler and Miller, 2005; Alsan and Goldin, 2019; Ferrie and Troesken, 2008; Anderson et al., 2019).[5] While there has long been agreement on the considerable importance of water filtration in explaining the mortality decline during the early 20th century, recent evidence has suggested that these results may have been overstated (Cutler and Miller, 2005; Anderson et al., 2019). Although research on water technologies and mortality has attracted considerable interest, little attention has been devoted to the role of water filtration in defeating cholera, a waterborne disease that caused millions of deaths during the 19th century and remains a feared disease in developing countries.[6]

To study whether filtered water protected against cholera, I use monthly house-level data on water contracts with individual-level mortality data from the city of Stockholm for the period 1860-1872. This unique historical setting combines partial access to filtered water within the city with a sizable citywide cholera epidemic in 1866. Using a difference-in-differences design, I test whether houses that already had access to filtered water before the shock had lower mortality during the shock than houses that did not.

I find that having access to filter water was protective during a cholera outbreak. The mortality rates of houses with filtered water and those without were similar in both levels and trends before and after the cholera outbreak. However, during

[4]This decrease in infectious disease mortality, termed the epidemiological transition, has been challenging for researchers to explain. The second half of the 19th century was a time coinciding with large-scale constructions of urban waterworks, real income increases, and an improved understanding of infectious diseases and how they spread. Separating these different explanations for declines in mortality in infectious diseases has been high on the agenda of researchers for many decades (Kesztenbaum and Rosenthal, 2017; Preston, 1975; Mckeown et al., 1975).

[5]Earlier historical evidence is summarized in a review by van Poppel and van der Heijden (1997), more recent historical evidence is provided by Helgertz and Önnerfors (2019); Peltola and Saaritsa (2019); Floris and Staub (2019), and research from developing countries on clean water and health includes, but is not limited to, Zhang (2012); Bhalotra and Venkataramani (2013); Galiani et al. (2009); Kota et al. (2015); Duflo et al. (2015); Watson (2006).

[6]Clearly, John Snow implicitly addressed this question in his seminal works on cholera in the UK. Following his mapping of cholera cases, there has been a series of replications of the initial findings. See, e.g., Coleman (2019) and references therein.

the shock, the mortality rates diverged significantly. Houses without filtered water experienced mortality rates that were up to twice as high as houses that already had filtered water. This finding stresses the historical importance of water filtration on mortality during time periods earlier than those previously been studied.[7] In later time periods, when cholera was absent in Europe and North America, it is possible that water filtration had a much smaller impact on mortality.

2 Background

During the 1850s, Stockholm experienced mortality rates well above 3 percent on average. Stockholm had the highest infant mortality rate in the country and had twice the national average mortality rate. During this time, four out of ten children born in Stockholm did not survive to see their first birthday. Life expectancy at birth followed the high infant mortality rate and was approximately 20 for men and 26 for women (Lindberg, 1980). This was a time when cholera was common during the summer months, and especially violent outbreaks hit the city in 1853 and 1857.

The poor disease environment had many explanations. Stockholm was the largest city in Sweden and expanded rapidly during the 1860s. Poverty was high, sanitation poor, and housing construction did not keep pace with the growing population. Poverty is evident in the high rate of out-of-wedlock births: 37 percent of all children in Stockholm were born out of wedlock. Observers at the time assigned much of the high infant mortality rate in Stockholm to high fertility rates among unmarried women (Berg, 1869). Although urban life at this time was nasty, brutish, and short, discussions about a water cleaning and citywide piped distribution system hinted of healthier times to come.

2.1 Filtered Water in Stockholm

Before the introduction of filtered water in Stockholm, its inhabitants relied on wells and nearby lakes for their daily drinking and washing water supply. There

[7] Troesken (2008) analyze time-series data from Chicago for the years 1850-1925 which include this earlier period. The authors do not specifically investigate cholera mortality.

were more than 300 private and 27 public wells at the time, but the water quality and reliability during summer was generally poor (Cronström, 1986).[8] The public wells were investigated by the local health commission in 1867, and only 6 out of 27 wells were deemed moderately suitable to drink from; the rest were recommended against drinking from.[9] Bacteriological analysis of the drinking water in Stockholm was not systematically undertaken until 1884 (Cronström, 1986).

The poor water quality could in part have been a result of inadequate sewerage infrastructure. Sewage irrigation in Stockholm consisted mainly of street ditches, some of which were covered. This allowed for ground infiltration and the released of waste into surrounding lakes. In many of the larger cities of continental Europe, underground self-contained sewerage systems had already been constructed in the 1840s. However, in Stockholm, it not be until 1875 that a large-scale system was constructed.[10]

Fiscal reasons delayed sewerage construction in Stockholm and generated resistance to the water cleaning and distribution system. In 1853, a plan for a water system with full city coverage was presented (Hansen, 1897). Local authorities were initially hesitant but subsequently encouraged to make the investment as cholera struck Stockholm during the 1850s. While the relationship between cholera and drinking water was unknown to almost everyone at the time, contemporary writers related disease to fumes emanating from garbage and fecal waste on the streets, the miasma theory of disease, which could be cleaned using pressurized water (Cronström, 1986).

In 1858, a modified version of the same plan received final approval, and the

[8] The wealthy inhabitants could buy water shipped from springs or wells outside the city where the water quality was much better, but most inhabitants had to rely on water from the wells.

[9] These results were based primarily on the levels of organic compounds and water hardness, the metrics used at this time to assess water quality. The commission urged the public not to use well water for drinking but to rely on the piped water now in place in large parts of the city. Information on this assessment was retrieved from the city archive in Stockholm.

[10] Piped water was seen as a prerequisite because a continuous water flow was need to prevent clogging of the sewers. A few public sewerages were built during the 1860s, and in 1872, there were four main lines. Locally, these could service a few percent of the population at most (Cronström, 1986). In 1875, the responsibility for city planning and construction was assigned to a newly created institution, the construction office (Byggnadskontoret), that had more access to the resources needed than its predecessor, Tredje Drätselnämnden. Citywide plans to build sewerage irrigation systems could then be approved. In 1876, it had become mandatory for houses to connect to a sewerage pipe if possible.

Stockholm Water Company was created. A few years later, in June 1861, the first 30 km of pipes were completed, and in October of the same year, for the first time, piped water was available to the inhabitants of Stockholm. The most central parts of Stockholm were the first with access to piped water.[11] These were wealthier parts of Stockholm, but the population was to a large extent mixed. Different income groups often shared the same building, with the wealthier living on the lower floors (Cronström, 1986). By 1872, the distribution network had already increased to 80 km in length, and to increase availability, a few public fountains were set up where water could be collected free of charge.

When access to a main pipe was available, house owners had to decide whether to finance an in-house service connection. The initial cost per room was 2.6 riksdaler (riksmynt) per year ($16 USD in current prices or 154 SEK). A shared tap on the bottom floor could be used by the tenants in each building, and there was no additional cost of utilization. The fixed cost of installing a tap, connected to the water supply by a service pipe, was on average 120 riksdaler ($754 USD in current prices), a considerable sum at the time for a working class household but rather modest if shared by many tenants.[12]

After some initial skepticism, water access became popular, and more houses were connected. In 1872, it was estimated that 70 percent of the population in Stockholm had in-house access to piped water, and water consumption increased during the studied time period (see figure 1).[13] In 1872, daily per user consumption was approximately 60 liters, which can be compared to approximately 180 liters today.[14]

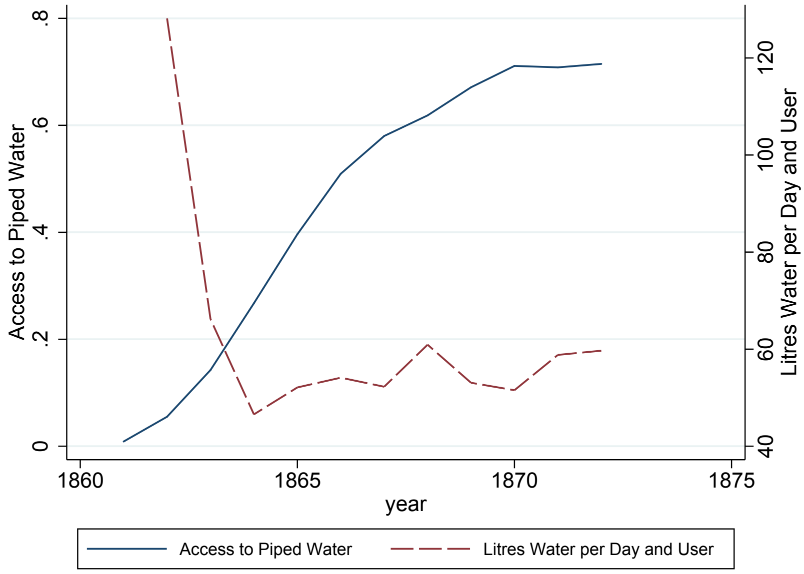
[11] Main pipes had been laid north, crossing Södermalm and continuing north through the old town (the small island in the middle) and over to Norrmalm. Figure 7 shows the initial network, constructed before 1861 (in blue), which is heavily concentrated in the old town and the southern parts of Norrmalm.

[12] See Stockholm city council propositions 1867 : 14 pp.5, link to web page. A male rural servant could expect to be receive approximately $1900 USD (2017) in compensation per year, including in-kind benefits (280 riksdaler rmt (1865)) (Edvinsson and Söderberg, 2011; Statistiska Central-byrån, 1868).

[13] Any city using lake water at this time, including Stockholm, had problems with water temperatures during the summer months. Reports from the time note that piped water was not popular to drink during the warmest time of the year as it became warm (Lindman, 1911, p. 274). The temperature would naturally vary within Stockholm depending on the time spent underground.

[14] See: http://www.drickkranvatten.se/virtuellt-vatten. Since there were no measurements at this time, some of this water went to industry, street cleaning, and cattle. However,

Figure 1: Take-up of Piped Water and Consumption



Note: This graph displays the uptake of piped water among households and the per user consumption of piped water. Access to piped water is calculated as the estimated number of household users divided by the total population. At this time, there was no water metering, so the measure of water usage is total consumption divided by the number of household users. This overestimates usage to some extent, as street cleaning and businesses also contributed to usage. Source: Kommunalförvaltning (1875).

2.2 The Technology for Cleaning and Distributing Water

The water cleaning plant took water from a nearby lake. At the time it was considered to provide adequate water quality and offered technical benefits (i.e., proximity to the water cleaning plant). Before deciding on the location of the water intake, the lake was examined in several places for organic compounds, where the closest location obtained the best results and was the strongest candidate. Later, as new sources of water were considered, more thorough measures of the water intake location and the supplied water were undertaken. The piped water was examined in 1879, and in terms of fecal residue, it was found to be superior to the piped water in London (Cronström, 1986).

The technology for cleaning water by slow filtration was discovered in the early 1800s, and the first plant for cleaning piped water was built in London in 1829

income statements from the Stockholm Water Company suggest that most of the water was used by households. In 1870, 70 percent of revenue from water consumption came from households. Industry and public buildings had water meters and paid for each unit of water used (Kommunalförvaltning, 1875).

(Huisman and Wood, 1974). The water cleaning facilities in Stockholm were also inspired by the newest plants in the UK.[15] The sand layers in the pools were seven feet deep with eight different layers of sand and stone, from fine sand to a layer of stones the size of coconuts (Cronström, 1986).[16] This technology is still viewed as an effective way of cleaning water of pathogens (Huisman and Wood, 1974).

While there are reasons to believe that the cleaning technology was effective, the piped distribution system was more problematic. For the main distribution net, pipes were made of cast iron coated in tar to prevent oxidation. The service pipes in buildings were made of lead, which were easy to shape in ways that made access around existing infrastructure easy. At this time, there were concerns about the health effects of lead, but these concerns were dismissed, as experimentation with other materials had shown that there existed no functional alternative (Cronström, 1986).[17]

2.3 Cholera – "The Blue Death"

Cholera is a highly contagious and lethal bacterial disease originating from the Ganges delta in India. Although records indicate that cholera has existed for a very long time, pandemics are not known before the early 19th century. During the 19th and early 20th centuries, cholera claimed millions of lives in recurring pandemics spreading throughout nearly the entire populated world (see table 1). Stockholm suffered severe outbreaks of cholera in 1834 and 1853, with deaths occurring in the thousands (Zacke, 1971). In Western Europe, there were few outbreaks of cholera after 1875.

One reason that cholera is such a feared disease comes from its rapid spread and high mortality rate among those infected and developing symptoms. The expression,

[15] Large pools were built to facilitate slow filtration in adequate amounts for the population of Stockholm. The plant was built at Skanstull in the southern part of Stockholm and initially had three slow filtration pools with a total area of 1,600 square meters.

[16] The top layer of fine sand was partially removed and washed around every sixth day during summer and more rarely during winter. When the top layer was at a minimum, new clean sand was added. The complete sand filter was replaced every other year (Hansen, 1897).

[17] Lead pipes are a concern for research on general health effects of filtered water because they could counteract any positive effect from clean water. However, studying the effect of filtered water on a brief cholera shock makes the pipe material less of a concern.

Table 1: Cholera Mortality Estimates for a Selected Sample of Countries

Cholera Pandemic	**Years**	**Countries Affected**	**Mortality**
First	1817-1824	Mainly in Asia	?
Second	1826-1837	Russia	100,000
		France	100,000
		Sweden	12,600
		Hungary	100,000
		UK	55,000
Third	1846-1860	Russia	1,000,000
		UK	52,000
		US	150,000
		Spain	236,000
		Sweden	20,000
Fourth	1863-1875	Russia	200,000
		Hungary	30,000
		Germany	115,000
		UK	14,300
		Netherlands	20,000
		Belgium	30,000
		Italy	130,000
		Sweden	4,500
Fifth	1881-1896	Germany (Hamburg)	8,600
		Russia	800,000
		Spain	60,000
		Italy	5,000
		France	5,000
Sixth	1899-1923	Russia	500,000

Note: These mortality numbers can be found primarily in Pollitzer (1954). Most numbers are crude estimates and inherently unreliable. There are many affected European countries where there are no mortality estimates at all. Especially for the earliest pandemics, information is lacking and highly uncertain. Nevertheless, it provides a broad picture of the disease spread in Europe. Many other countries in the world, e.g., Japan and India, were severely affected by cholera but are not described here.

"The Blue Death," comes from the blue skin tone of cholera-infected individuals suffering from extreme dehydration. Infected humans can produce up to 20 liters of diarrhea a day, rapidly leading to acute dehydration and a lethal imbalance of electrolytes. During the 19th century, mortality from cholera was often as high as 50 percent (Koch, 1894). At present, with access to oral and intravenous rehydration therapies, mortality is much lower and mainly affects small children who are more sensitive to dehydration (UNICEF, 2019).

Although the British physician John Snow already had a good idea about how cholera spread among people in 1849, this was not common knowledge in Sweden, or elsewhere, for several decades (Snow, 1849). In a pamphlet distributed by the physicians association in Stockholm during the 1866 cholera outbreak, one can read how to prevent being exposed to the disease.[18] People were encouraged to eat and drink in moderation, to drink less liquor, not to go out in the morning without having eaten anything, keep their feet warm and dry and bring fresh air indoors by opening windows. Drinking water was not mentioned at all.

Robert Koch, one of the most important contributors to bacteriological science, confirmed the ideas of John Snow when he isolated the bacterium Vibrio cholerae in 1884.[19] In addition to isolating bacteria, Koch also observed and documented epidemic outbreaks of infectious diseases. While working with a cholera outbreak in Hamburg in 1892, he noted that residents of Altona, a separate municipality within the city of Hamburg, had almost no cholera cases, while Hamburg had many (see figure 2). The only difference between the two municipalities, right at the border, was that Altona had its own water supply taken downstream from the river Elbe, while Hamburg took its water upstream. As Hamburg was a large city with its sewer released into the Elbe, Altona should have been worse off, although flood tides reversed the flow of the river regularly. However, Altona had used a water

[18]This pamphlet can be found at the Swedish historical web page Stockholmskällan (SLL, 1866).

[19]The medical community was resistant to these new ideas, which competed with the ruling miasma theory of disease, but the evidence was indisputable. The change in perspectives and the success of the germ theory of disease was consolidated in 1905 when Koch received the Nobel prize in Medicine and Physiology for discovering the cause of another bacterial disease, tuberculosis. Furthermore, Koch was not the first to observe and attribute the disease of cholera to a bacillus. The Italian Filippo Pacini had already done this in 1854 and published a report that was completely ignored by the medical community for decades (UCLA, 2020).

filter since 1859 to clean their drinking water, while Hamburg did not. He wrote,

"On both sides of the frontier [between Hamburg and Altona], the state of the soil, the buildings, the sewerage, the population, in short all the conditions the are important in this connection, are perfectly similar, and yet the cholera in Hamburg spread only to the frontier of Altona, and stopped there." (Koch, 1894, pp. 25)

Figure 2: Cholera Cases in Hamburg and Altona at the City Border 1892-1893

Note: Dots represent cholera cases, and the thick line represents the border between Hamburg (below) and Altona (above). Altona had its water filtered, while Hamburg did not. Both took their drinking water from the river Elbe, bottom-left corner. Reproduced from Exner (2015) describing the cholera outbreak in Hamburg.

Clearly, the difference in water supply between Hamburg and Altona was crucial for the spread of cholera at that time. However, whether it was due to water filtration or some other unobserved differences in water quality generated by differences in water intake has not been established (BMJ, 1893).

2.4 Cholera in Stockholm 1866

The last severe outbreak of cholera in Stockholm was in 1866. The first and last recorded deaths from cholera took place on June 29 and October 29 (Sundhets-collegii, 1869, pp.25). The cholera epidemic was unexpected in Stockholm. The first documented cases were discovered in both Stockholm and the second largest city in Sweden, Gothenburg, at almost the same time. During these few months, cholera was widespread in the city. All eight geographical parishes comprising Stockholm reported cholera cases and fatalities.[20]

During the cholera outbreak in 1866, most of the deceased were prime-age adults and small children. The vast majority of cholera deaths, and cases, were concentrated in adults between 20 and 50 years of age (see the left panel of figure 3). By this with the right panel of the same figure, which depicts the age distribution of all deaths before 1866, we can see that prime-age adults were disproportionately affected by cholera compared to, for example, small children.

That adults were infected more frequently, and subsequently died from cholera, is strengthened by other information on cholera infections reported at the time (Sundhets-collegii, 1869, pp.25). Adults more often contracted cholera but died less often. Children under age 10 died in almost half of all documented infections, i.e., a case fatality rate (CFR) from cholera of approximately 0.42. Prime-aged adults (aged 20-50) had a much lower CFR of 0.26, while the elderly, over age 60, had an even higher CFR of 0.55. This suggests that cholera infections were most common in adults between 20 and 50 years of age.

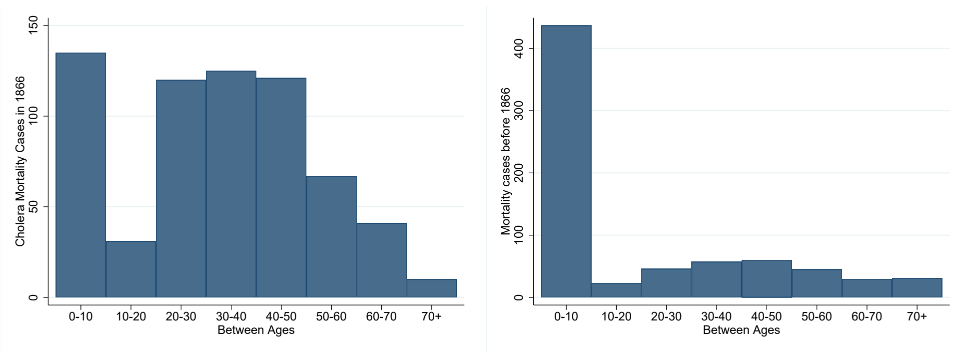
At present, small children are at much higher risk of succumbing to cholera than adults. Cholera mortality among children under five makes up approximately half of all cases in endemic countries (UNICEF, 2019). Historical evidence, however, shows a different picture. Davenport et al. (2019) found that in London, despite the high severity of the cholera outbreaks in 1849, 1854, and 1866, infant mortality was largely unaffected during these outbreaks. Historical evidence from Denmark shows that the cholera mortality rate during the 1850s was lowest among children under 5 (2.7 %) and highest among the elderly (16.7 %) (Phelps et al., 2017).

Unfortunately, the social class of those who died of cholera in 1866 is not well documented. However, as different social classes often lived in the same buildings and in the absence of knowledge on how cholera spreads, the disease could have been difficult to avoid. Detailed information from an earlier cholera outbreak in Stockholm shows that excess mortality was 2-3 times the average mortality among

[20]The parishes reported between 20 and 122 fatalities from cholera during the shock. In total, there were 655 documented deceased and 2,200 infected.

the upper classes, making up some 30 % of the population. For the "rest" of the population, mortality was 3-4 times the average during the outbreak (Zacke, 1971, pp. 167). As with most infectious diseases, the poor are normally more susceptible to infections and subsequent death than other social classes. Nevertheless, while poverty was a risk factor for dying from cholera, everyone was at risk.

Figure 3: Age Distribution of Cholera Deaths in 1866 and Previous Years



Note: The left panel shows the age distribution for documented cholera cases in Stockholm according to official records. Source data for the left panel can be found at Sundhets-collegii (1869, pp.25). The right panel describes the age distribution of all-cause mortality cases in the data averaged over the period 1860-1865 and only using deaths from July through October.

3 The Historical Data Sources

To study how clean water affects cholera mortality, data from several historical sources have purposely been collected and digitized. Data from three different sources are used: the 1860 Swedish census, archival water contract information, and mortality data from parish registers. These data sources are detailed below.

Since the treatment is clustered at the house level, possible confounding factors should be addressed at this level. To obtain more information on the buildings under study and their populations, I digitized part of the 1860 census in Sweden concerning Stockholm. These data contain information on existing addresses of residential buildings in 1860, which I use to set up a population of houses. In addition to defining residential buildings, the data contain information on the number of residents at each address, the number of households in the building, and the sex composition at the time of the census. These are important characteristics related to the mortality and socioeconomic status of residents.

From the mortality data, I exclude deaths at public and private institutions, such as poor houses, prisons, orphanages, military installations, and hospitals. These institutions rarely have valid addresses (a street name and street number combination), and they always contain large populations. To match this selection in the census data, I exclude households with more than 50 inhabitants in 1860 (average household size is approximately 5 in the sample). Establishments housing the poor and military installations often contain up to 100-300 inhabitants per household (as defined in the census data). After excluding buildings without a valid address, removing institutions reduces the sample by approximately 56 households. In total, the census data contain 3,201 valid addresses with residents in 1860.[21]

Information on filtered water was retrieved from water contract lists preserved at the Stockholm City Archives for the public water company. This archive includes a contract book dating from 1861 to 1872, which is the source of the contract data. The contract data contain information on the day, month, and year of each contract, as well as the address, parish, block, and number of rooms debited (see figure 8).

The water data contain 3,530 contracts, of which I can match 2,700 to the population of houses described above. Those that cannot be matched either belong to buildings that had no residential population (e.g., a factory or warehouse), to houses that were constructed after 1860, or to houses that were missing from the census for other reasons. Of these matched contracts, 1,564 are unique. As houses changed owners, a new contract had to be written and thereby counted twice or more in the data. Furthermore, I am not able to separate addresses with the same street name and street number, but separated by letters (e.g., a, b, c, d). These addresses were part of the same building but with different entry doors and are unfortunately not well distinguished in the census data. I merge these special cases at their street name and street number.

Mortality data are available from the Swedish genealogical society (Sveriges släktforskarförbund). Recently, they digitized all deaths in Sweden between 1860

[21]However, it is likely that the available census data are incomplete. Summing up the total number of inhabitants in the data, without exclusions, I have 107,191 residents compared to the official number of 117,000. Some parts of the archive could unfortunately have gone missing over the years.

and 2017 in their searchable database, The Swedish Death Index (SSF, 2018). This database contains information on date of birth, date of death, sex, marital status at time of death, and address at the time of death. The primary underlying source for this database is the parish records. I merge the mortality data with the census population and water contract data using month and address of death.

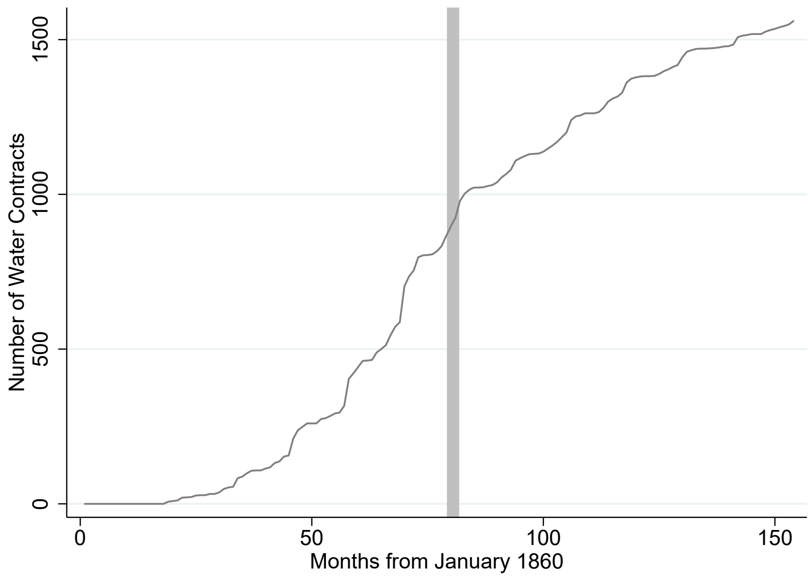
The data do not contain information on cause of death. Although this is a limitation, using all-cause mortality allows me to capture the effect on cases where cholera is an underlying, but not the primary, cause of death and on cases where cause of death is missclassified. The mortality data further allow me to separately construct mortality measures at different ages and by sex.

Observing a consistent population of residential houses over time implies that some of the mortality (and water contract) data are lost. The crude mortality data, between 1860 and 1872, contain 52,172 entries of disease. A meaningful fraction of these belong to institutions that are not used in the data. At least 8,000 entries belong to an institution, where the largest one is an orphanage (Allmäna Barnhuset) that contains approximately 3,000 deaths. A further 9,000 entries have a valid address but cannot be matched to the sample. This is due to houses missing in the census data, newly constructed residential buildings not existing in 1860, or buildings that were not inhabited in 1860 but were later. Finally, I have approximately 35,000 deaths that are within the sample selection criterion, of which 6,500 are missing on either street address, street number, or both. The final sample contains 28,557 valid observations of diseased individuals. More information on the matching procedure is provided in the Appendix (see section 7).

4 Empirical Approach

To analyze how access to filtered water affected cholera mortality, I use July-October as months with cholera and define a dummy variable indicating these months in 1866 ($Cholera_t$). I further divide the sample into two parts, buildings that had water in June 1866 and those that did not, to have a treatment measure that cannot respond to the shock. Since the shock was only active for four months, this narrow time frame reduces the scope for having an in-house tap installed in response to the shock. Figure 4 shows that only a small fraction of the water users connected during the cholera shock of 1866, more reflecting it being summer when the temperature allowed for digging than a demand response.

Figure 4: Cumulative Number of Water Contracts in the Data



Note: The figure shows cumulative unique contracts in Stockholm between 1860 and 1872. Unique means that only the first occurrence of a contract at a specific address is counted. The same building could have several contracts in the data if the house, e.g., changed owners during the period. The cholera shock is marked in gray.

Since the aim of this analysis is to compare mortality between groups of houses during a cholera shock, it is vital to understand which houses belong to these different groups. The treated group are houses that had connected to the water network and signed a contract prior to June 1866. The control group consists of houses that did not have water at that time. Table 2 presents descriptive statistics of the variables used in the analysis. At the time of the cholera epidemic of 1866, 26 percent of the houses in the data had in-house filtered water. Access had increased to 49 percent as of the end of 1872.

Moreover, the houses that connected to the water distribution network are relatively large, especially early adopters. The average number of inhabitants in these houses is 38, compared to 29 in the control group. While houses with and without water in 1866 had a similar mix of men and women, there are somewhat more

Table 2: Descriptive Statistics

	Mean	SD	Min	Max	Observations
Water Variables					
In-house Water	0.239	0.426	0	1	499356
Water Before Cholera	0.260	0.438	0	1	499356
Water Before Cholera x Cholera	0.007	0.081	0	1	499356
Water Before 1873	0.490	0.500	0	1	499356
Months From 1859	79	45	1	156	499356
Rooms (Ever Water Sample)					
Rooms per House	33	32	1	312	1567
Residents per room	1.79	1.88	0.02	24.00	1567
Census Data					
Popualtion in 1860	31.46	28.06	1	346	3201
Fraction female 1860	0.57	0.13	0	1	3201
Households per House 1860	6.30	5.42	1	62	3201
Streets in data	23	21	1	137	3201
City Blocks in Data	250	142	1	492	3201
Addresses in data	1601	924	1	3201	3201
Mortality Variables (Monthly)					
Mortality Count	0.057	0.275	0	34	499356
Mortality (per pop)	0.002	0.015	0	1	499356
Mortality (100k pop)	215	1520	0	100000	499356
Female Mortality (100k pop)	107	1087	0	100000	499356
Male Mortality (100k pop)	108	1018	0	100000	499356
Child Mortality (100k pop)	97	996	0	100000	499356

Note: The water and mortality variables come at a monthly frequency at the address level. The "Ever water sample" includes addresses that eventually obtained a water contract up to 1872. The census provides cross-sectional data for 1860 only.

households in the treated houses, reflecting that the latter were relatively larger.[22] This can to a large extent be explained by the initial distribution network being

[22]This can be more directly seen by comparing houses that had water before 1866 to those that had water after 1866 but before 1873. Houses with water before 1866 have on average 39 rooms, while later adopters only have 25 on average. See table 5.

concentrated in the central part of Stockholm, where there were also large houses. I include these control variables in the main regression model to account for these differences.

As the main outcome variable, I use the mortality rate per 100,000 population at each address and month. To create this variable, I collapse the data to the address and month level. I divide by the population in 1860 and multiply by 100,000. The same procedure is used for measures of child mortality and for mortality by sex.

The empirical approach is a difference-in-differences design, following the early work of John Snow (Lechner, 2011). To this end, I create an interaction between *water before the cholera shock* and *the four cholera months*. With time and address fixed effects, the main effects of the interaction are subsumed. I include a time- and address-varying water dummy in all specifications ($Water_{pt}$) that describes the timing of water contracts at each address, leading to the following model:

$$Mortality_{pt} = \beta \cdot (Water_p^{t=1866(June)} \cdot Cholera_t) + \theta \cdot Water_{pt} + \mathbf{X'}\mathbf{t}_t + \delta_t + \gamma_p + \upsilon_{pt}. \quad (1)$$

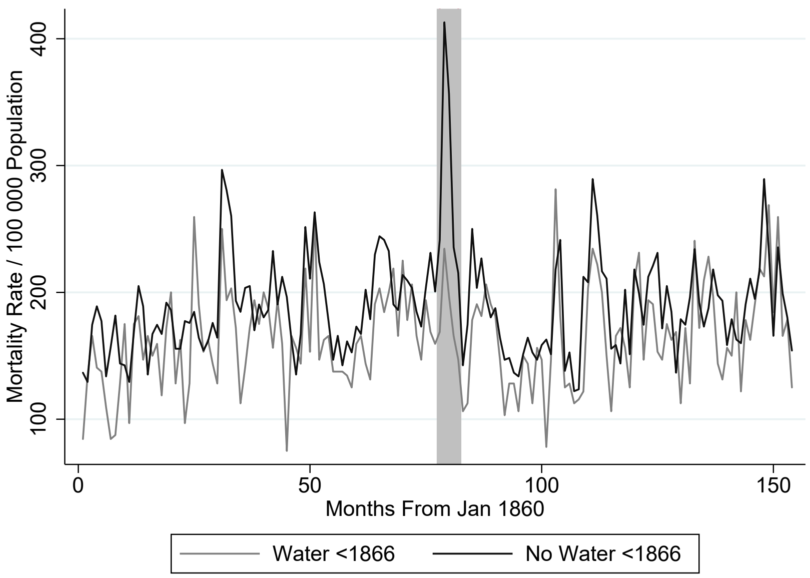
In equation 1, γ_p and δ_t are street address and unique month fixed effects. $\mathbf{X'}\mathbf{t}_t$ are the predetermined house characteristics described above that interact with the time dummies, i.e., population, fraction female, and number of households. υ_{pt} is a random disturbance term. With this design, I am comparing the difference in mortality during the cholera shock between houses with and without water to the difference in mortality between the same groups before and after the cholera shock.

The regression model, equation 1 above, is estimated using weighted least squares (WLS) while weighting by the number of residents in each building in 1860. Since there are large differences in house populations and mortality at the address-month level is uncommon, weighting can reduce noise in the data and thereby increase precision. Standard errors are clustered at both the time dimension (month) and at the city block level to account for common shocks in time and geographically correlated persistence over time for houses close to one another (Cameron and Miller, 2015).

5 The Effect of Having Filtered Water during a Cholera Outbreak

The first evidence I present is graphical and shows the monthly mortality rate over time (per 100,000 population in 1860) for houses with and without filtered water in June 1866. These results are presented in figure 5. The water group, containing approximately 25 percent of the houses in the data, is naturally somewhat more volatile than the group without water in 1866. Nevertheless, there is a clear co-movement between the two series, in both trends and levels. If these groups were different on characteristics related to mortality, we would expect to see at least differences in mortality in levels.

Figure 5: All-Age Mortality Rate between Water and No Water in 1866

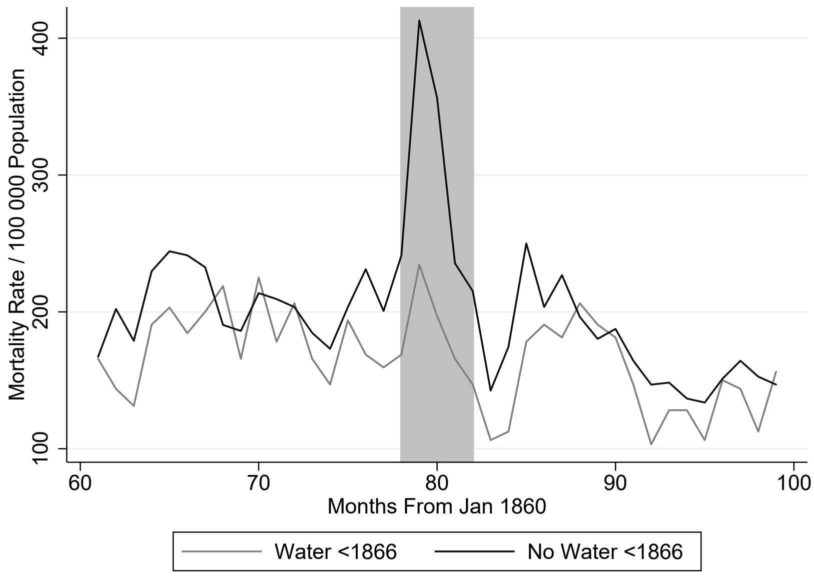


Note: The cholera shock is marked in gray. The two groups described in the figure are defined based on having a water contract in June 1866. The water group (black) had a water contract before June 1866, while the no-water group (gray) did not.

The exception to this co-movement is during the time of the cholera shock. During those months, there is a clear spike in mortality in the group of houses without filtered water, while there is almost no response in houses with filtered water. After the shock, the series converge again and move closely together. Consistent with other accounts of cholera, the disease spreads rapidly after entering the population and rather quickly decreases in intensity (Koch, 1894). According to official records, peak mortality was only four weeks after the first case was recorded (Sundhets-collegii, 1869, pp.25).

The difference in mortality as cholera spread in Stockholm is even more apparent if we focus on the 1.5 years before and after the shock. Figure 6 shows this restricted version, where we can see more clearly that mortality peaked in the first month after the start of the epidemic. Furthermore, mortality was only higher for houses without water until the second month into the epidemic and then subsumed in intensity to more comparable levels. Mortality before cholera followed the same trend between the two groups and returned to a common path after cholera. After the epidemic, there was a small dip in mortality for both treated and controls. This dip could reflect both behavioral changes after the epidemic and a shift in mortality over time.[23]

Figure 6: All Age Mortality Rate between Water and No Water in 1866



Note: The cholera shock is marked in gray. The two groups described in the figure are defined based on having a water contract in June 1866. The water group (black) had a water contract before June 1866, while the no-water group (gray) did not.

To be more specific about the magnitude of the protective effect of filtered

[23]Behavioral changes could include people being extra cautious in terms of movements or interactions after the epidemic. Shifting mortality over time means that people who in the absence of the epidemic would have died after the shock, but now died during the shock.

water, regression models are used that estimate equation 1. Table 3 (panel A) shows separate estimates for females, males, and for both sexes under the age of five. In column 1, the interaction effect of filtered water during the cholera shock is -131 (se= 60) deaths per 100,000 population. Since the average monthly mortality rate was 259 per 100,000 population during the cholera period, the estimated effect amounts to a 50 % reduction in mortality over the four months of cholera. The effect is almost exactly the same for males and females and somewhat smaller for children under five, likely reflecting that young children were not infected by cholera to the same extent as adults.

Table 3: Does Filtered Water Protect Against Cholera?

	Outcome: Mortality Rate per 100,000 Population			
	All	Women	Men	Under 5
	(1)	(2)	(3)	(4)
Panel A.				
Water x Cholera	-131.31972	-66.69967	-64.62005	-28.78169
	(59.60666)	(30.08416)	(33.21549)	(20.11003)
Panel B.				
Water x Cholera:				
First Month	-207.15492	-122.31294	-84.84198	-29.09534
	(59.82604)	(27.84917)	(32.56129)	(17.18516)
Second Month	-185.65061	-69.23194	-116.41867	-69.11705
	(60.76466)	(26.96423)	(33.79723)	(20.12995)
Third Month	-76.76850	-65.67822	-11.09028	-26.32293
	(53.19313)	(28.05164)	(23.29181)	(17.04716)
Fourth Month	-55.93768	-9.72467	-46.21302	9.33404
	(26.08427)	(8.11208)	(17.98883)	(4.07899)
Mean	215.2020	112.8707	115.3932	102.5151
Observations	499356	499356	499356	499356

Note: Each column in panel A displays results from a separate regression for the interaction term: water before July 1866 and the four months of cholera in 1866 (July-October). Panel B reports four interactions for each of the cholera months. All models include month and address fixed effects and controls for the number of households, population size, and fraction of females residing in the building in 1860 (interacted with time effects). In total, the data contain 3201 unique addresses. Standard errors are clustered at both the time level ($G_1 = 154$) and at the city block level ($G_2 = 492$). All regressions are WLS using house populations in 1860 as weights.

Panel B displays the same models but where I define treatment for each of the four cholera months. This provides more information on the dynamics of the epidemic over time. It is clear from panel B that cholera was most active during the first two months and had a much smaller impact thereafter. During the first month, houses with piped water had 207 fewer deaths per 100,000 population. During the last cholera month, there were only 56 fewer deaths in houses with filtered water. This trend is similar for men, women, and children.

Although the regression results are consistent with the visual evidence presented in the figures above, the estimated effects could still be sensitive to deviations from the specification described in equation 1. Particularly since the treatment time period is short and only constitutes a small fraction of all time periods in the data, the estimates could be sensitive to alternative specifications. Table 4 presents regression estimates from several different specifications, highlighting the robustness of the main findings presented above.

Column 1 of table 4, shows the population-unweighted difference-in-differences estimate without control variables. First, excluding control variables is relevant to make clear how important they are for identification. The point estimate is reduced to -85 (se= 36) without control variables but remains significant and is not statistically different from the WLS estimate with control variables. Second, weighting the observations by population size can be important if treatment effects are heterogeneous in that dimension. The same model without control variables is also estimated using WLS, with weights according to the 1860 population, resulting in a slight increase of the point estimate (estimated effect: -97 (se= 38)). This congruence between weighted and unweighted estimates serves as a specification check and indicates that there is limited heterogeneity in the effect across houses with different population sizes (Solon et al., 2015).

Another concern is that the underlying (mortality) data are in actual counts at this very low level of aggregation and that further include many zero observations (the mortality count mean is 0.0572). Since it is not clear that OLS performs well under this type of data generating process, I compare OLS estimates on count data to a count data estimator (i.e., a Poisson estimator). When estimating the effects of filtered water using these two estimators, I find that they are very close. OLS suggests that filtered water reduced cholera mortality by 46 percent (at the mean of the outcome), while the Poisson estimator finds that mortality was reduced by approximately 44 percent.[24]

Table 4: Specification Checks

	Mortality Rate per 100'000 Pop		Mortality Count	
Model:	OLS	WLS	OLS	Poisson
	(1)	(2)	(3)	(4)
Water x Cholera	-85.91743	-97.09771	-0.02631	-0.37175
	(35.70893)	(37.98707)	(0.01011)	(0.13571)
Mean	215.2020	215.2020	0.0572	0.0572
Observations	499356	499356	499356	456300

Note: Each column displays results from a separate regression for the interaction term: water before July 1866 and the four months of cholera in 1866 (July-October). The model used includes month and address fixed effects. There are no control variables in these models to provide comparisons to other specifications with control variables. In total, the data contain 3201 unique addresses. Standard errors are clustered at both the time level ($G_1 = 154$) and at the city block level ($G_2 = 492$). In column 4, standard errors are clustered at the block level only.

6 Conclusion

Cholera emerged as a feared and lethal disease in European cities during the first half of the 19th century. However, long before effective treatment or a vaccine against cholera was invented, cholera mortality in continental Europe had declined dramatically. This development coincides with water improvements in many European cities beginning in the 1830s. This paper studied the case of Stockholm and how access to filtered water on tap protected against cholera during its last outbreak in 1866. To do so, new and detailed information on mortality and in-house filtered tap water at the address level during the 1860s was used.

The findings showed that filtered water was an important technology for preventing cholera-related mortality. Residents of houses with access to tap water were nearly immune to cholera, in contrast to those without such access. These

[24] Using the correction for a log difference approximation $(e^{0.372} - 1) * 100 \approx 44\%$.

findings emphasize the importance of waterworks during the 19th century in reducing mortality from cholera. If water filtration was most effective in mitigating truly waterborne diseases, such as cholera, it may have been less important for the mortality decline at later stages of development when cholera was not present anymore (Anderson et al., 2019).

Water filtration is not the only possible reason for the decline in cholera, which could have been aided by other public health efforts. Better quarantine rules for migrants and merchants, improvements in living standards (nutrition), and medical advances (rehydration therapy) could have reduced morbidity and mortality from cholera and other epidemic waterborne diseases (Preston, 1975; Mckeown et al., 1975). However, the analysis showed that filtered water on its own was sufficient to almost entirely prevent cholera mortality among those with access.

Although this research found that filtered water was effective in preventing cholera mortality, quantifying the full benefit of waterworks in relation to cholera is complicated due to the positive externalities of almost all contagious disease-mitigating technologies. Constructing waterworks in a specific city could, e.g., reduce the probability of cholera appearing in other, nearby cities. In this way, waterworks in cities situated at the center of extensive trade networks could have been important for stopping the further spread of cholera. While heavily affected by imported infectious diseases, these cities were likely early adopters of technologies that reduced disease propagation, such as waterworks or strict quarantine rules.

An interesting avenue for future research would be to further this line of reasoning by mapping out trade routes, cholera outbreaks, and waterworks. If important trade hubs constructed water cleaning infrastructure, other more peripheral cities could have benefited from the reduced spread of disease. In this way, it could potentially be possible to quantify some of the positive externalities from waterworks.

References

Alsan, M. and Goldin, C. (2019). Watersheds in child mortality: The role of effective water and sewerage infrastructure, 1880–1920. *Journal of Political Economy*, 127(2):586–638.

Anderson, D. M., Charles, K. K., and Rees, D. I. (2019). Re-examining the contribution of public health efforts to the decline in urban mortality. *American Economic Journal: Applied Economics*, Z(X):YY.

Berg, F. (1869). Om dödligheten i första lefnadsåret. *Statistisk tidskrift*, 23:435–493.

Bhalotra, S. R. and Venkataramani, A. (2013). Cognitive development and infectious disease: Gender differences in investments and outcomes. IZA Discussion Paper 7833, IZA, Bonn.

BMJ (1893). The cholera of 1892 in hamburg. *The British Medical Journal*, 1(1677):373–375.

Cameron, C. A. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372.

Coleman, T. (2019). Causality in the time of cholera: John snow as a prototype for causal inference. Working paper, University of Chicago.

Cronström, A. (1986). *Stockholms tekniska historia: Vattenförsörjning och avlopp*. Liber, Almqvist & Wiksell, Uppsala.

Cutler, D. and Miller, G. (2005). The role of public health improvements in health advances: The twentieth-century united states. *Demography*, 42(1):pp. 1–22.

Davenport, R. J., Satchell, M., and Shaw-Taylor, L. M. W. (2019). Cholera as a 'sanitary test' of british cities, 1831–1866. *The History of the Family*, 24(2):404–438.

Duflo, E., Greenstone, M., Guiteras, R., and Clasen, T. (2015). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation. Working Paper 21521, National Bureau of Economic Research.

Edvinsson, R. and Söderberg, J. (2011). A consumer price index for sweden 1290-2008. *Review of Income and Wealth*, 57:270–292.

Ewald, P. W. (1991). Waterborne transmission and the evolution of virulence among gastrointestinal bacteria. *Epidemiology and Infection*, 106(1):83–119.

Exner, M. (2015). Cholera epidemic in hamburg, germany 1892. In Bartram, J., editor, *Routledge Handbook of Water and Health*, pages 644–647. Routledge, London and New York.

Ferrie, J. P. and Troesken, W. (2008). Water and chicago's mortality transition, 1850–1925. *Explorations in Economic History*, 45(1):1 – 16.

Floris, J. and Staub, K. (2019). Water, sanitation and mortality in swiss towns in the context of urban renewal in the late nineteenth century. *The History of the Family*, 24(2):249–276.

Galiani, S., Gonzalez-Rozada, M., and Schargrodsky, E. (2009). Water expansions in shantytowns: Health and savings. *Economica*, 76(304):607–622.

Hansen, F. V. (1897). Stockholms vattenledning. In Dahlgren, E. W., editor, *Stockholm : Sveriges hufvudstad skildrad med anledning af allmänna konst- och industriutställningen 1897*, pages 322–352. Beckmanska Boktryckeriet, Stockholm.

Helgertz, J. and Önnerfors, M. (2019). Public water and sewerage investments and the urban mortality decline in sweden 1875–1930. *The History of the Family*, 24(2):307–338.

Huisman, L. and Wood, W. (1974). Slow sand filtration. Technical report, World Health Organization.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press.

Kesztenbaum, L. and Rosenthal, J.-L. (2017). Sewers' diffusion and the decline of mortality: The case of paris, 1880–1914. *Journal of Urban Economics*, 98(Supplement C):174 – 186. Urbanization in Developing Countries: Past and Present.

Knutsson, D. (2017). Water improvement and health: Historical evidence on the effect of filtering water on urban mortality. Research papers in economics 2017:2, Stockholm University.

Koch, R. (1894). *Professor Koch on the bacteriological diagnosis of Cholera, Water-filtration and Cholera, and the Cholera in Germany during the winter of 1892-1893*. William R. Jenkins, New York.

Kommunalförvaltning, S. (1875). *Berättelse angående Stockholms kommunalförvaltning : jämte statistiska uppgifter för samma och föregående tid*. Beckmanska Boktryckeriet, Stockholm.

Kota, O., Shinichiro, S., and Genya, K. (2015). Public health improvements and mortality in interwar tokyo: A bayesian disease mapping approach. Discussion paper 2015-06, Tokyo Institute of Technology.

Lechner, M. (2011). The estimation of causal effects by difference-in-difference methods. *Foundations and Trends(R) in Econometrics*, 4(3):165–224.

Lindberg, F. (1980). *Växande stad. Stockholms stadsfullmäktige 1862-1900.* Liber, Almqvist & Wiksell, Uppsala.

Lindman, C. (1911). *Sundhets och befolkningsförhållanden i sveriges städer 1851-1909*. Schmidts Boktryckeri, Hälsingborg.

MacKinnon, J. G. and Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254.

McKeown, T. and Brown, R. G. (1955). Medical evidence related to english population changes in the eighteenth century. *Population Studies*, 9(2):119–141.

Mckeown, T., Record, R. G., and Turner, R. D. (1975). An interpretation of the decline of mortality in england and wales during the twentieth century. *Population Studies*, 29(3):391–422. PMID: 11630508.

Peltola, J. and Saaritsa, S. (2019). Later, smaller, better? water infrastructure and infant mortality in finnish cities and towns, 1870–1938. *The History of the Family*, 24(2):277–306.

Phelps, M., Perner, M. L., Pitzer, V. E., Andreasen, V., Jensen, P. K. M., and Simonsen, L. (2017). Cholera Epidemics of the Past Offer New Insights Into an Old Enemy. *The Journal of Infectious Diseases*, 217(4):641–649.

Pollitzer, R. (1954). History of the disease. In *Cholera Studies*, number 10, pages 421–461. WHO.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, 29(2):231–248.

SLL (1866). Underrättelse för allmänheten. Svenska Läkare-Sällskapet. Accessed: 2020-05-26. https://stockholmskallan.stockholm.se/ContentFiles/SSA/Biblioteket/124_13_42_kolera1866.pdf.

Snow, J. (1849). *On the mode of communication of Cholera*. Wilson and Ogilvy.

Solon, G., Haider, S. J., and Wooldridge, J. M. (2015). What are we weighting for? *J. Human Resources*, 50:301–316.

SSF (2018). Sveriges dödbok 1860-2017. Technical report, Sveriges Släktforskarförbund. DVD, Searchable database.

Statistiska Central-byrån, S. (1868). *Jordbruk och boskapsskötsel. Hushållningssällskapens berättelser*. P.A. Norsted och Söner, Stockholm. Bidrag till Sveriges Officiella Statistik.

Sundhets-collegii (1869). *Hälso och Sjukvården 1*. P.A. Norsted och Söner, Stockholm. Bidrag till Sveriges Officiella Statistik.

Troesken, W. (2008). Lead water pipes and infant mortality at the turn of the twentieth century. *The Journal of Human Resources*, 43(3):553–575.

UCLA (2020). Who first discovered vibrio cholerae? Department of Epidemiology. Accessed: 20200301. https://www.ph.ucla.edu/epi/snow/firstdiscoveredcholera.html.

UNICEF (2019). Cholera. Accesses 2019-12-12: https://www.unicef.org/cholera/.

van Poppel, F. and van der Heijden, C. (1997). The effects of water supply on infant and childhood mortality: a review of historical evidence. *Health Transition Review*, 7(2):113–148.

Watson, T. (2006). Public health investments and the infant mortality gap: Evidence from federal sanitation interventions on u.s. indian reservations. *Journal of Public Economics*, 90(8–9):1537 – 1560.

Westlund, B. and Stahre, N.-G. (1986). *Stockholms gatunamn : innerstaden*. Liber/Allmänna förl., Stockholm, 1. uppl. edition.

Zacke, B. (1971). *Koleraepidemin i Stockholm 1834*. PhD thesis, Stockholms Universitet, Stockholm. Monografier utgivna av Stockholms Kommunalförvaltning 32.

Zhang, J. (2012). The impact of water quality on health: Evidence from the drinking water infrastructure program in rural china. *Journal of Health Economics*, 31(1):122 – 134.

7 Appendix

7.1 Data set construction

The mortality data used in this paper were collected and transcribed by the Swedish genealogical society using parish church books for all deceased individuals in Sweden from 1860 onward. The information available in these records is sex, address at death, parish of death, parish of birth, date of birth/death, and marital status (including widowed). Loading the raw data, I have 52,172 recorded deaths for the city of Stockholm between 1860 and 1872. I do not use mortality recorded at institutions or organizations such as elderly homes, hospitals, orphanages, or delivery clinics. Although I could map out some of these institutions to addresses, many of these would have a huge impact on the estimates, and changes in the institutional environment, new institutions or locations, would make the analysis rather complicated. Furthermore, there are many missing addresses for these institutions, and some of them keep their own parish records (death books).

Below, I provide some information on where the missing mortality data come from. First, there are 2,520 deaths that are completely missing address information. A total of 543 of these are noted as dying at a delivery clinic (and a handful of these are also at a hospital). At the time, delivery clinics (by law) allowed parents to remain anonymous and hand over newborns to foster care (organized by orphanages). These infants would not be exposed to the home environment of the parents and could be placed in foster care anywhere in Sweden. A further 350 of these deaths are recorded at military installations, some of which kept their own death records (but often lacking the address of death). Additionally, more than 300 records missing address information come from parishes/congregations not belonging to the Swedish protestant church. The English, Finnish, and Jewish congregations in Stockholm would also gather members from other parts of Sweden without a congregation. A total of 1,250 of the 2,520 observations without address information are listed in the geographical parishes of Stockholm and should be in the data if a useful address had been provided. These are unidentified deaths and deaths of individuals without a known address. Many of these have double zeros as entries for month and day of birth, suggesting that their age has been estimated. This could mean that they were visiting or recently migrants to Stockholm not yet registered as residents.

Next, of the 49,652 observations that are left, 11,979 lack a proper address. Almost 3,000 deaths without a proper address come from the death books of the largest orphanage in Stockholm at the time. Many more of these are institutional deaths, and many of them should not be in the sample; 6,094 belong to the 8 geographical parishes and could possibly be in the data. Of these 6,094 deaths, more than half of the observations still belong to an institution (as described in the address field). Notably to an orphanage (Stora Barnhuset, 510 observations), the city morgue (Stockholms stads Bårhus, 376 observations), and work houses for the poor (Fattighuset, 817 observations). Out of these nearly 12,000 deaths, at least 8,000 are listed in an institutional death book or have an institution written in the address field.

In the final stage, the different data sources were matched together. The census was used to set up a sample of addresses that were populated and existed in 1860. The water data, containing clearly written contracts, were rather easy to merge. The mortality data were more complicated. Since many streets changed names during the 1850s, there were many names not merely misspelled or misinterpreted but also describing a street using an old name. I looked through all street names that I could not easily match to the census using internet resources (mainly Stockholmskällan providing historical photos and documents from Stockholm) and Westlund and Stahre (1986) where historical street names in the inner city are detailed. Furthermore, I made use of a map from 1863 where all existing street names were listed (unfortunately, the street numbers were not written out). With this map, I could benchmark that a street number existed. I used this map and inspected all street names that were given a street code.[25]

Using these sources, I was able to match most of the address fields to a street name. Approximately 6,500 deaths with a valid street name and number could not be matched to the census. These missing data could be due to typos, transcription

[25]See Stockholm map 1863 at Stockholmskällan, last accessed 2020-03-09.

errors, or addresses not existing in 1860. If there were several streets with the same name, of which there were a handful in Stockholm at this time, I only matched those where the geographical parish matched the street name. For example, "Lilla Badstugatan" exists in both the Sofia and Clara parishes. Only entries with that address registered in either of these geographical parishes were used to avoid placing deaths at the wrong address. See figure 9 for an example of the matching. The matching sheet is available from the author upon request.

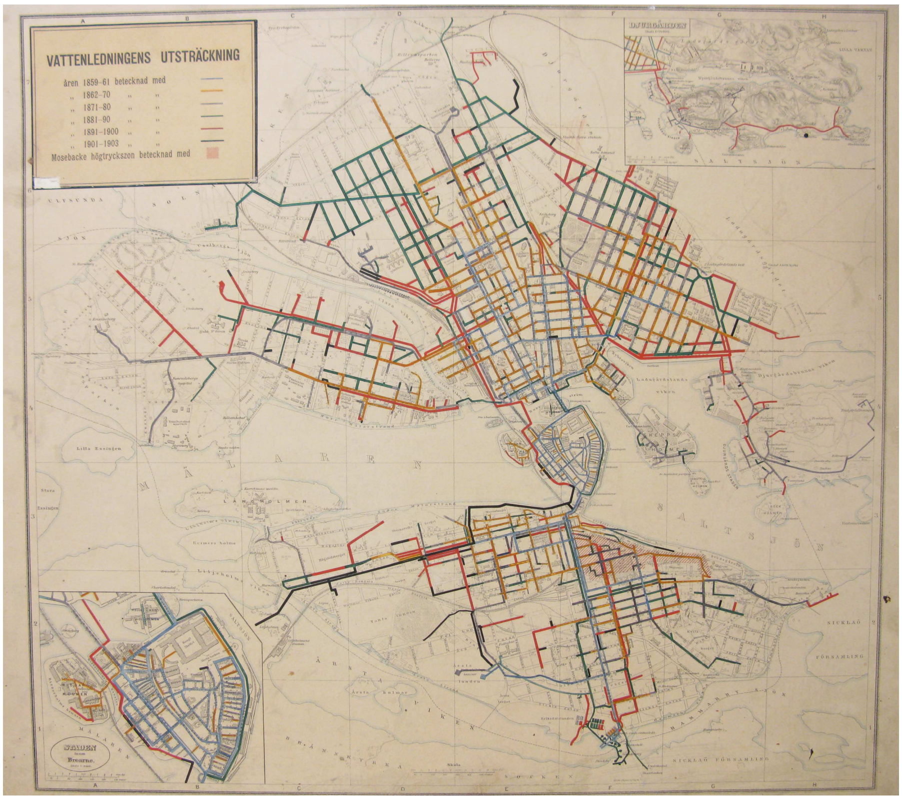
7.2 Additional Figures and Tables

Table 5: Treated, Controls, and Ever Treated

Variable (mean)	All	Treated	Control	Control | Water After Cholera
Population	32	38	29	36
Fraction Female	0.57	0.57	0.56	0.57
Households per House	6.3	7.3	6.0	7.3
Rooms per House	.	39	.	25
Residents per Room	.	1.46	.	2.16

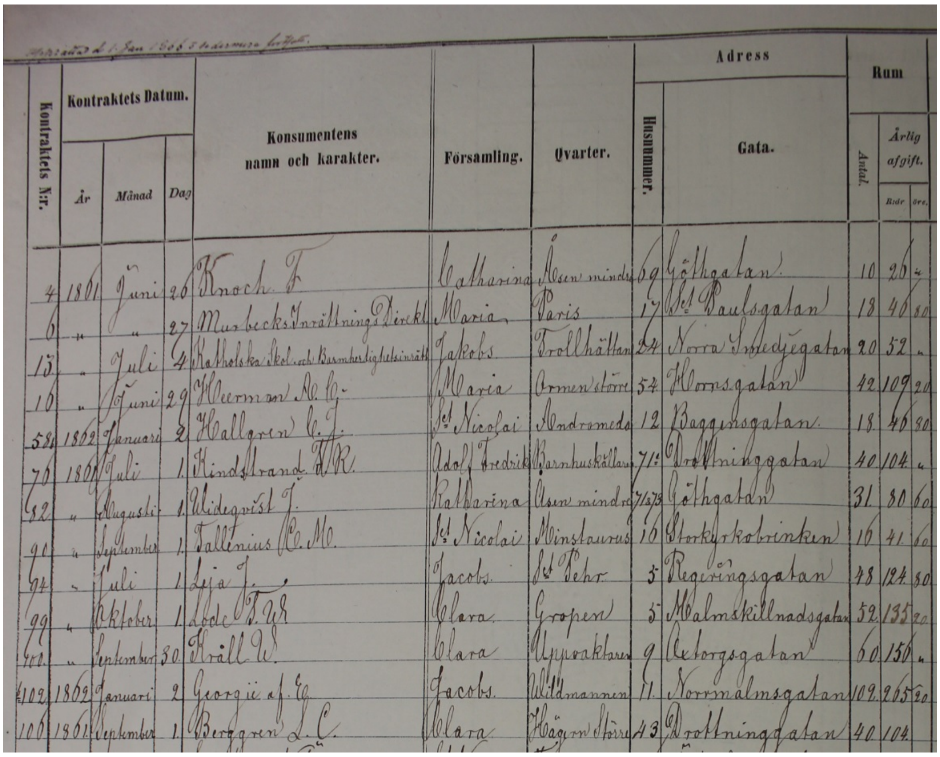
Note: The table shows averages of different variables measured in 1860 or recorded in the water contracts (rooms in building). These averages are presented for different groups in the data. Ever treated are those that eventually had a water contract in the data, i.e., before 1873. The ever treated controls are then houses that were listed for a contract after June 1866 but before 1873.

Figure 7: Map of Water Network Construction over Time



Note: Blue lines represent the distribution network at the time of its introduction in 1861. Yellow lines show the additional pipes laid during the 1860s, and the other colors represent later construction. Source: Stockholm Water Company archive at Stockholm City Archives.

Figure 8: Water Contract List Data



Kontraktets N:r	Kontraktets Datum. År	Månad	Dag	Konsumentens namn och karakter.	Församling.	Qvarter.	Husnummer.	Adress Gata.	Rum Antal.	Årlig afgift. Rdr	Öre.
4	1861	Juni	26	Knoch F.	Catharina	Åsen mindre	69	Göthgatan	10	26	"
6	"	"	27	Musbecks Inrättnings Direkt[illegible]	Maria	Paris	17	St Paulsgatan	18	46	80
13	"	Juli	4	Katholska Skol och Barmhertighetsinrättn.	Jakobs	Trollhättan	24	Norra Smedjegatan	20	52	"
16	"	Juni	29	Heerman [illegible]	Maria	Ormen större	54	Hornsgatan	42	109	20
58	1862	Januari	2	Hallgren C. J.	St Nicolai	Andromeda	12	Bagginsgatan	18	46	80
76	1861	Juli	1	Lindstrand [illegible]	Adolf Fredrik	Barnhuskällaren	71	Drottninggatan	40	104	"
82	"	Augusti	1	Widegvist J.	Katharina	Åsen mindre	71&73	Göthgatan	31	80	60
91	"	September	1	Tallenius [illegible]	St Nicolai	Minotaurus	16	Storkyrkobrinken	16	41	60
94	"	Juli	1	Leja J.	Jacobs	St Pehr	5	Regeringsgatan	48	124	80
99	"	Oktober	1	Lodé [illegible]	Clara	Gropen	5	Malmskillnadsgatan	52	135	20
101	"	September	30	Kråll W.	Clara	Uppvaktaren	9	Oxtorgsgatan	60	156	"
102	1862	Januari	2	Georgii af G.	Jacobs	Wildmannen	11	Norrmalmsgatan	109	265	20
106	1861	September	1	Berggren L. C.	Clara	Hägern Större	43	Drottninggatan	40	104	

Note: Example of the contract list data between 1861 and 1872. From the left, the data include contract number, contracting date, contract holder name, resident parish, block name, street address and number of rooms debited (total yearly cost). Source: Stockholm Water Company archive at Stockholm City Archives.

Figure 9: Example of Matching

	A	B	C	D	E	F
1	sname	scode	Forsaml	Församling	water	Dead
2	?-gatan			Adolf Fredrik		1
3	Kungsholmsbrogatan			Klara		1
4	. Vattugatan			Klara		1
5	.. Kungsholmsbrogata			Klara		1
6	Adolf Fredrik kyrkogata	58		Adolf Fredrik		1
7	Adolf Fredrik Södra Gränd	58		Adolf Fredrik		1
8	Adolf Fredrik Södra Kyrkogata	58		Adolf Fredrik		1
9	Adolf Fredriks Kyrkogata	58		Adolf Fredrik		1
10	Adolf Fredriks S Kyrkogata	58	5	Adolf Fredrik	1	
11	Adolf Fredriks Södra Gata	58		Adolf Fredrik		1
12	Adolf Fredriks Södra Kyrkogata	58	5	Adolf Fredrik		
13	Adolf Fredriks Södra Kyrkogata	58		Adolf Fredrik		1
14	Adolf Fredriks torg	600		Maria		
15	Adolf Fredrikskyrkogata	58		Adolf Fredrik		1
16	Akademiegränd	154	6	Klara	1	
17	Akademigränd	154	6	Klara		
18	Akademigränd	154		Klara		1
19	Albano	601		Johannes		
20	Allm gränd	602		Stockholms finska		1
21	Allmäna gränd	602		Hedvig Eleonora		1

Note: Example of how the matching took place. Unique street names from all three data sources were appended to one file and sorted on name. Sname is the street name, providing a common street code for all spellings. Scode is the street code that I assigned to all contracts, houses, and deaths that with the same name. Forsaml and Församling are the parishes from the death data and contract list data. Water and dead columns indicate the source of each street name, meaning that if no indication on these variables, the source is the census. A missing scode means that no match was done, and the data were excluded.